To control for or not to control for?

Concepts of Mediation Analysis in Epidemiology

Javier Sanchez – Ian Dohoo

Centre for Veterinary Epidemiological Research

UNIVERSITY of PRINCE EDWARD ISLAND

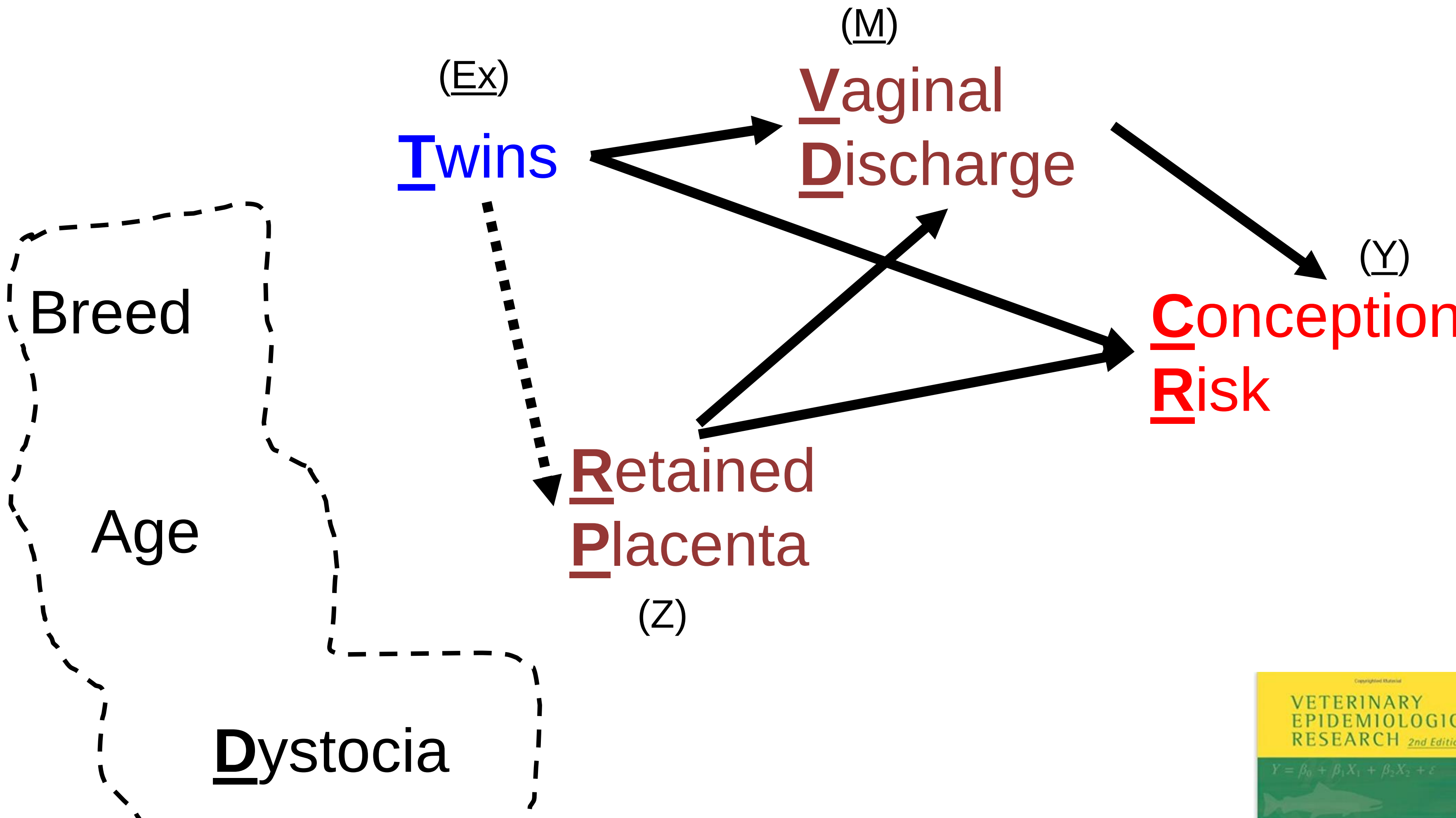
Outline

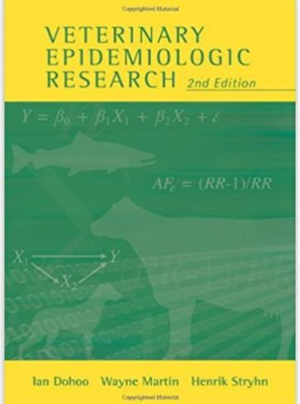
- Why mediation analysis?
- Example - Reproductive diseases in cattle
- Approaches and biases
- Software
- Remarks

Why mediation analysis?

- Need to understand different pathways that could explain **<u>Ex</u>**posure->Outcome (**<u>Y</u>**) relationship
- Typically interest is in **total effect Ex->Y**
  - do not control for intervening variables!??
- We could also estimate an indirect effect
  - mediating (intervening) variable
- …or the direct effect
  - effect not explained by the mediator (intervening) variable

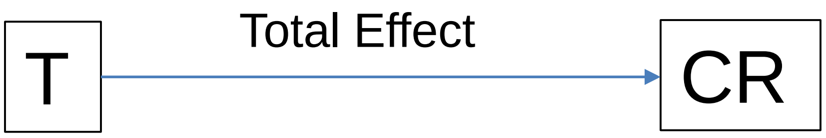
Reproductive Diseases in Dairy Cattle



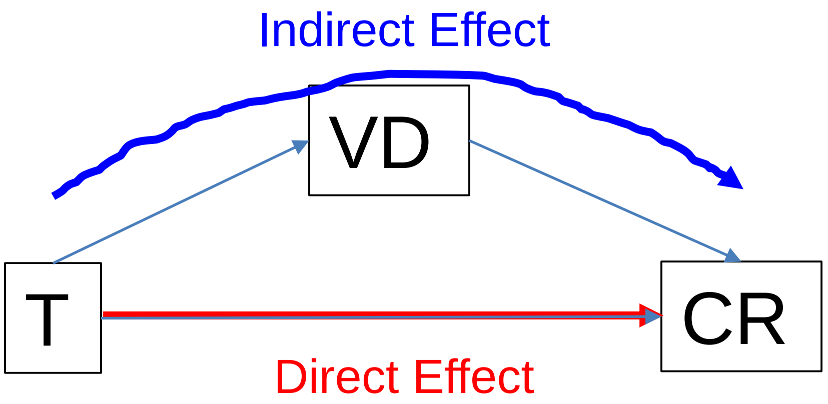


Mediation analysis – Example

- Effect of Twin ("T") on 1st Service Conception Risk ("CR")



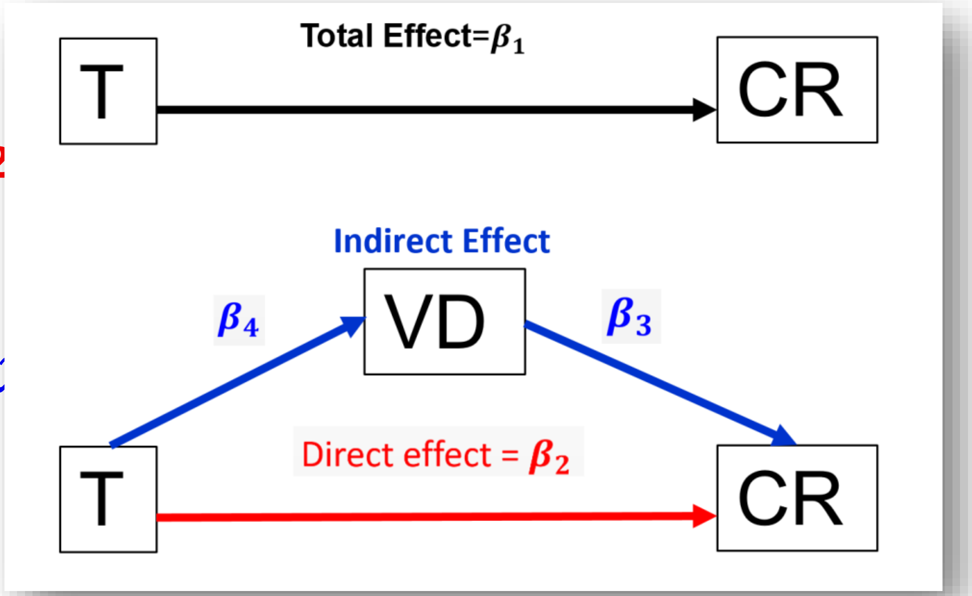
- Role of Vaginal Discharge ("VD") as potential mediator (*indirect/direct effects*)



CVER

Mediation analysis – Traditional approach

- Baron and Kenny (1986) paper
- Model for outcome
  - E[Y|ex]= $\beta_0 + \boldsymbol{\beta_1} * Ex\ \ (without\ mediator)$
    - Total effect = $\boldsymbol{\beta_1}$
  - $E[Y|ex, m] = \beta_0 + \boldsymbol{\beta_2} * Ex + \beta_3 * M$ (with mediator)
    - Direct effect = $\boldsymbol{\beta_2}$
    - Indirect effect = $\boldsymbol{\beta_1} - \boldsymbol{\beta_2}$
- Model for mediator
  - E[M|ex] = $\beta_0\ + \beta_4 * E$
    - Indirect effect = $\boldsymbol{\beta_4} * \boldsymbol{\beta_3}$



Mediation analysis – Traditional approach

- Non linear models (eg. logistic) mediation effects doesn’t correspond with causal effects
  - Eg. **Total Effect ≠ Direct Effect + Indirect Effect**
  - Causal interpretation?
    - counterfactual model

- Biases
  - Incorrect statistical design
  - Three main type of biases

CVER

Counterfactual framework

Cow	T	VD	CR
Daisy	Yes	No	Yes
Betsy	No	Yes	Yes

Counterfactual framework

Cow	Herd	Parity	T	VD	CR
Daisy	2	4	Yes	No	Yes
Betsy	3	6	No	Yes	Yes

Counterfactual framework

	Cow	Herd	Parity	Twin	Vag. Disch	CR	
	Daisy	2	4	Yes	No	Yes	
	Daisy	2	4	No	No	?	counterfactual
	Betsy	3	6	No	Yes	Yes	
	Betsy	3	6	Yes	Yes	?	counterfactual

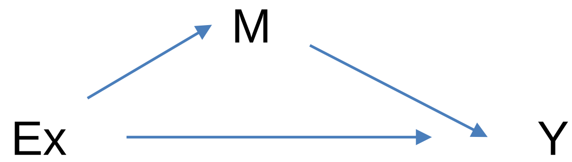
Counterfactual framework – No Mediator

- Compare: $Y_{Ex=1}$ {Y(1)} to $Y_{Ex=0}$ {Y(0)}
  - For any individual – only Y(1) or Y(0) is observed
  - $E(Y_i(1) - Y_i(0)) = E(Y_i|Ex=1) - E(Y_i|Ex=0)$
  - "difference in means estimator"
  - Unbiased estimate of Average Causal Effect (ACE)

CVER

Counterfactual framework - Mediator

- Mediation



  - Potential outcome $Y_i(ex, m)$
  - Many possible – only one observed for each individual

- Natural indirect effect (NIE)
  - NIE=E[Y(1,M(1)]-E[Y(1, M(0)]
  - compares Y under M = M(Ex=0) vs M(Ex=1)
  - changes in Y if Ex is fixed at (ex) but M changes by amount expected from changing Ex from 0 to 1

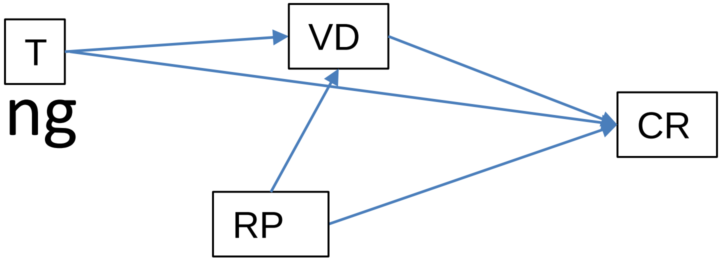
Counterfactual framework - Mediator

- Natural direct effect (NDE)
  - NDE=E[Y(1,M(ex)]-E[Y(0, M(ex)]
  - compares Y under Ex=1 vs Ex=0, fixing M=M(Ex=ex)
  - changes in Y if M is fixed at level corresponding to (ex) but Ex changes from 1 to 0

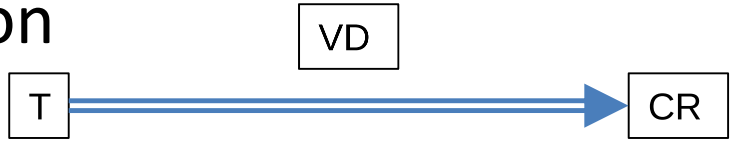
- Total causal effect – decomposed as:
  - TCE=E[Y(1)]-E[Y(0)] = NDE + NIE

Mediation analysis – Biases

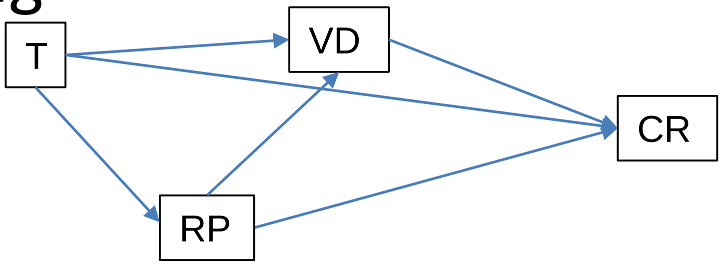
- Three main sources of bias
  - 1) Mediator-Outcome Confounding



  - 2) Exposure-Mediator Interaction

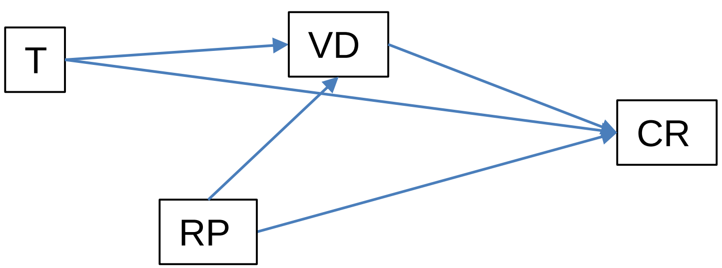


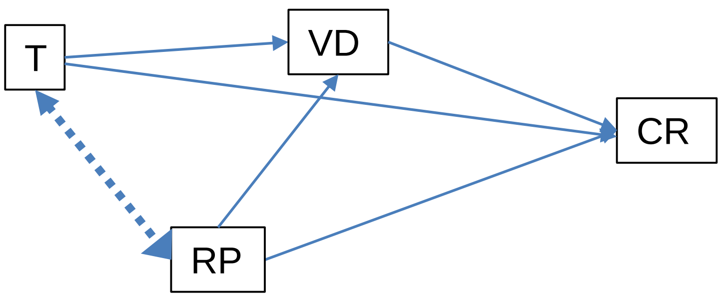
  - 3) Mediator-Outcome Confounding affected by Exposure



Mediator-Outcome Confounding

- Conditioning on VD (mediator) creates spurious association between RP (confounder) and T (exposure)





  - M-bias
  - VD – collider variable

- Assume no other confounders are present

Commands – Stata / R

- Stata
  - medeff command – simulation based
    - The Stata Journal 11, 605-619. Causal mediation analysis.
  - Syntax

Mediator model

Outcome model

```
medeff (logit vag_disch twin rp)
       (logit cr twin vag_disch),
mediate(vag_disch)
treat(twin) sims(1574)
```

- R
  - medflex and mediation packages

Commands – Stata / R

- Effects binary outcome
  - natural indirect effect

$$OR_1^{NIE} = \frac{P(Y_{1M_1} = 1)/P(Y_{1M_1} = 0)}{P(Y_{1M_0} = 1)/P(Y_{1M_0} = 0)}$$

  - Natural direct effect

$$OR_0^{NDE} = \frac{P(Y_{1M_0} = 1)/P(Y_{1M_0} = 0)}{P(Y_{0M_0} = 1)/P(Y_{0M_0} = 0)}$$

  - Total effect

$$OR^{TE} = OR^{NIE} \times OR^{NDE}$$

Mediation – Outcome confounder - medeff

- Stata

MEDIATOR MODEL

Twin calving cows are ~ 5 times more likely to have vaginal discharge

vag_disch	Coef.	Robu Std.			[95% Conf.	Interval]
twin	1.159373	.5263		8	.1277918	2.190954
rp	1.818323	.2539		0	1.320623	2.316023
_cons	-3.297454	.1422009	-23.19	0.000	-3.576163	-3.018746

OUTCOME MODEL

Twin calving cows are ~ 0.27 times more likely to get pregnant

cr	Coef.	Robus Std. E			[95% Conf.	Interval]
twin	-1.304015	.5452356	-2.39	0.017	-2.372657	-.2353728
vag_disch	-.3786963	.2507406	-1.51	0.131	-.8701387	.1127462
rp	-.3997027	.1905788	-2.10	0.036	-.7732303	-.0261751
_cons	-.2408816	.0542461	-4.44	0.000	-.347202	-.1345613

Effect	Mean	[95% Conf.	Interval]
Average Mediation	.0056289	-.0018744	.0220223
Average Direct Effect	.2429117	.0584458	.3576201
% of Tot Eff mediated	.021562	.0154744	.0773487

Mediation – Outcome confounder

- ACME: Average Causal Mediated Effect
  - Natural Indirect effect
  - The increase in Y brought on by increasing M by the amount that would result from changing Ex from 0 to 1, while holding Ex constant
- ADE: Average Direct Effect
  - The increase in Y brought about by changing Ex from 0 to 1 while holding M constant
- ACME + ADE: Total causal effect
  - Increase in Y brought about by changing Ex from 0 to 1, and allowing M to change correspondingly

Commands – Stata / R

- Model types

	Outcome Type					
	Continuous		Binary		Counts	
Mediator type	Stata	**R**	Stata	**R**	Stata	**R***
Continuous	Y	Y	Y	Y	N	Y
Binary	Y	Y	Y	Y	N	Y
Count	N	Y	N	Y	N	Y
Ordinal/Nominal	N	Y	N	Y	N	Y

* Poisson and Negative Binomial

CVER

Commands – Stata / R

- Biases

	Stata		R
Bias	medeff	G-formula*	medflex
1) M-Y confounder	Y	Y	Y
2) Ex-M interaction	Y	Y	Y
3) M-Y affected by Ex	N	Y	N

*see references

CVER

Summary

- Control for or not to control for?
  - depend on the question!
  - control - quantify (“unbiased”) causal direct and indirect pathways
    - implement/target interventions
- Rely on a biological plausible causal model
  - emphasize the use of a causal diagram!!
- Fast growing area of research
  - underused methods
  - complexity of the problem
  - understanding counterfactual framework

CVER

References

- Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology 51, 1173-1182.
- Richiardi, L., Bellocco, R., Zugna, D., 2013. Mediation analysis in epidemiology: methods, interpretation and bias. International Journal of Epidemiology 42, 1511–1519. doi:10.1093/ije/dyt127
- Petersen, M.L., Sinisi, S.E., van der Laan, M.J., 2006. Estimation of Direct Causal Effects: Epidemiology 17, 276–284. doi:10.1097/01.ede.0000208475.99429.2d
- Lange, T., Vansteelandt, S., Bekaert, M., 2012. A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. American Journal of Epidemiology 176, 190–195. doi:10.1093/aje/kwr525
- Hicks, R. and Tingley, D., 2011. Causal mediation analysis. The Stata Journal 11, 605-619.
- Rhian, D., De Stavola, B.L., Cousens, S.N., 2011. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula.
- Steen, J., Loeys, T., Moerkerke, B., Vansteelandt, S., n.d. Medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models.